

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Matalon, Reuben
 Kaul, Rajinder
 Gao, Guang Ping
 Balamurugan, Kuppareddi
 Michals-Matalon, Kimberlee

(ii) TITLE OF INVENTION: Aspartoacylase Gene, Protein, and
 Methods of Screening for Mutations Associated with Canavan
 Disease

(iii) NUMBER OF SEQUENCES: 27

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
 (B) STREET: 2200 Clarendon Boulevard, Suite 1400
 (C) CITY: Arlington
 (D) STATE: Virginia
 (E) COUNTRY: U.S.A.
 (F) ZIP: 22201

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/128,020
 (B) FILING DATE: 29-SEP-1993
 (C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Hamlet-King, Diana
 (B) REGISTRATION NUMBER: 33,302
 (C) REFERENCE/DOCKET NUMBER: Shutt 1

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 703-243-6333
 (B) TELEFAX: 703-243-6410
 (C) TELEX: 64191

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1435 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: 159..1097

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTGTAACAGA AAATTAAAT ATACTCCACT CAAGGGAATT CTGTACTTTG CCCTTTTGGT 60
 AAAGTCTCAT TTACATTCT AAACCTTCT TAAGAAATC GAATTCCTT TGATCTCTCT 120

00965807-100101

TCTGAATTGC AGAAATCAGA TAAAACTAC TTGGTGAA ATG ACT TCT TGT CAC	173
Met Thr Ser Cys His	5
ATT GCT GAA GAA CAT ATA CAA AAG GTT GCT ATC TTT GGA GGA ACC CAT	221
Ile Ala Glu Glu His Ile Gln Lys Val Ala Ile Phe Gly Gly Thr His	10 15 20
GGG AAT GAG CTA ACC GGA GTA TTT CTG GTT AAG CAT TGG CTA GAG AAT	269
Gly Asn Glu Leu Thr Gly Val Phe Leu Val Lys His Trp Leu Glu Asn	25 30 35
GGC GCT GAG ATT CAG AGA ACA GGG CTG GAG GTA AAA CCA TTT ATT ACT	317
Gly Ala Glu Ile Gln Arg Thr Gly Leu Glu Val Lys Pro Phe Ile Thr	40 45 50
AAC CCC AGA GCA GTG AAG AAG TGT ACC AGA TAT ATT GAC TGT GAC CTG	365
Asn Pro Arg Ala Val Lys Lys Cys Thr Arg Tyr Ile Asp Cys Asp Leu	55 60 65
AAT CGC ATT TTT GAC CTT GAA AAT CTT GGC AAA AAA ATG TCA GAA GAT	413
Asn Arg Ile Phe Asp Leu Glu Asn Leu Gly Lys Lys Met Ser Glu Asp	70 75 80 85
TTG CCA TAT GAA GTG AGA AGG GCT CAA GAA ATA AAT CAT TTA TTT GGT	461
Leu Pro Tyr Glu Val Arg Arg Ala Gln Glu Ile Asn His Leu Phe Gly	90 95 100
CCA AAA GAC AGT GAA GAT TCC TAT GAC ATT ATT TTT GAC CTT CAC AAC	509
Pro Lys Asp Ser Glu Asp Ser Tyr Asp Ile Ile Phe Asp Leu His Asn	105 110 115
ACC ACC TCT AAC ATG GGG TGC ACT CTT ATT CTT GAG GAT TCC AGG AAT	557
Thr Thr Ser Asn Met Gly Cys Thr Leu Ile Leu Glu Asp Ser Arg Asn	120 125 130
AAC TTT TTA ATT CAG ATG TTT CAT TAC ATT AAG ACT TCT CTG GCT CCA	605
Asn Phe Leu Ile Gln Met Phe His Tyr Ile Lys Thr Ser Leu Ala Pro	135 140 145
CTA CCC TGC TAC GTT TAT CTG ATT GAG CAT CCT TCC CTC AAA TAT GCG	653
Leu Pro Cys Tyr Val Tyr Leu Ile Glu His Pro Ser Leu Lys Tyr Ala	150 155 160 165
ACC ACT CGT TCC ATA GCC AAG TAT CCT GTG GGT ATA GAA GTT GGT CCT	701
Thr Thr Arg Ser Ile Ala Lys Tyr Pro Val Gly Ile Glu Val Gly Pro	170 175 180
CAG CCT CAA GGG GTT CTG AGA GCT GAT ATC TTG GAT CAA ATG AGA AAA	749
Gln Pro Gln Gly Val Leu Arg Ala Asp Ile Leu Asp Gln Met Arg Lys	185 190 195
ATG ATT AAA CAT GCT CTT GAT TTT ATA CAT CAT TTC AAT GAA GGA AAA	797
Met Ile Lys His Ala Leu Asp Phe Ile His His Phe Asn Glu Gly Lys	200 205 210
GAA TTT CCT CCC TGC GCC ATT GAG GTC TAT AAA ATT ATA GAG AAA GTT	845
Glu Phe Pro Pro Cys Ala Ile Glu Val Tyr Lys Ile Ile Glu Lys Val	215 220 225

25iii

GAT TAC CCC CGG GAT GAA AAT GGA GAA ATT GCT GCT ATC ATC CAT CCT Asp Tyr Pro Arg Asp Glu Asn Gly Glu Ile Ala Ala Ile Ile His Pro 230 235 240 245	893
AAT CTG CAG GAT CAA GAC TGG AAA CCA CTG CAT CCT GGG GAT CCC ATG Asn Leu Gln Asp Gln Asp Trp Lys Pro Leu His Pro Gly Asp Pro Met 250 255 260	941
TTT TTA ACT CTT GAT GGG AAG ACG ATC CCA CTG GGC GGA GAC TGT ACC Phe Leu Thr Leu Asp Gly Lys Thr Ile Pro Leu Gly Gly Asp Cys Thr 265 270 275	989
GTG TAC CCC GTG TTT GTG AAT GAG GCC GCA TAT TAC GAA AAG AAA GAA Val Tyr Pro Val Phe Val Asn Glu Ala Ala Tyr Tyr Glu Lys Lys Glu 280 285 290	1037
GCT TTT GCA AAG ACA ACT AAA CTA ACG CTC AAT GCA AAA AGT ATT CGC Ala Phe Ala Lys Thr Thr Lys Leu Thr Leu Asn Ala Lys Ser Ile Arg 295 300 305	1085
TGC TGT TTA CAT TAGAAATCAC TTCCAGCTTA CATCTTACAC GGTGTCTTAC Cys Cys Leu His 310	1137
AAATCTGCT AGTCTGTAAG CTCCTTAAGA GTAGGGTTGT GCCTTATTCA ACTGCATACA	1197
TAGCTCCTAG CACAGTGCCT TATTCGGTAG GCATCTAAGC AAATTTCTTA AATTAATTAA	1257
TATATCTTTA AAGATATCAT ATTTTATGTA TGTAGCTTAT TCAAAGAAGT GTTTCCTATT	1317
TCTATATAGT TTATTATACA TGATACTTGG GTAGCTCAAC ATTCTTAATA AACAGCCTTT	1377
GTATTCAGAA TATAAAATTG AAATAGATAT ATATAAAGTT AAAAAAAAAA AAAAAAAAAA	1435

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 313 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 83
- (D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 105
- (D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 108
- (D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 146
- (D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site

(B) LOCATION: 264
 (D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:
 (A) NAME/KEY: Modified-site
 (B) LOCATION: 117
 (D) OTHER INFORMATION: /note= "Potential N-glycosylation site"

(ix) FEATURE:
 (A) NAME/KEY: Active-site
 (B) LOCATION: 18..24
 (D) OTHER INFORMATION: /note= "Consensus sequence predicted to be involved in catalysis"

(ix) FEATURE:
 (A) NAME/KEY: Active-site
 (B) LOCATION: 275..278
 (D) OTHER INFORMATION: /note= "Consensus sequence predicted to be involved in catalysis"

(ix) FEATURE:
 (A) NAME/KEY: Active-site
 (B) LOCATION: 283..289
 (D) OTHER INFORMATION: /note= "Consensus sequence predicted to be involved in catalysis"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Thr	Ser	Cys	His	Ile	Ala	Glu	Glu	His	Ile	Gln	Lys	Val	Ala	Ile
1				5					10					15	
Phe	Gly	Gly	Thr	His	Gly	Asn	Glu	Leu	Thr	Gly	Val	Phe	Leu	Val	Lys
	20							25					30		
His	Trp	Leu	Glu	Asn	Gly	Ala	Glu	Ile	Gln	Arg	Thr	Gly	Leu	Glu	Val
	35						40					45			
Lys	Pro	Phe	Ile	Thr	Asn	Pro	Arg	Ala	Val	Lys	Lys	Cys	Thr	Arg	Tyr
	50					55				60					
Ile	Asp	Cys	Asp	Leu	Asn	Arg	Ile	Phe	Asp	Leu	Glu	Asn	Leu	Gly	Lys
65					70					75				80	
Lys	Met	Ser	Glu	Asp	Leu	Pro	Tyr	Glu	Val	Arg	Arg	Ala	Gln	Glu	Ile
			85					90						95	
Asn	His	Leu	Phe	Gly	Pro	Lys	Asp	Ser	Glu	Asp	Ser	Tyr	Asp	Ile	Ile
		100						105					110		
Phe	Asp	Leu	His	Asn	Thr	Thr	Ser	Asn	Met	Gly	Cys	Thr	Leu	Ile	Leu
	115						120					125			
Glu	Asp	Ser	Arg	Asn	Asn	Phe	Leu	Ile	Gln	Met	Phe	His	Tyr	Ile	Lys
	130					135					140				
Thr	Ser	Leu	Ala	Pro	Leu	Pro	Cys	Tyr	Val	Tyr	Leu	Ile	Glu	His	Pro
145					150					155					160
Ser	Leu	Lys	Tyr	Ala	Thr	Thr	Arg	Ser	Ile	Ala	Lys	Tyr	Pro	Val	Gly
			165						170					175	
Ile	Glu	Val	Gly	Pro	Gln	Pro	Gln	Gly	Val	Leu	Arg	Ala	Asp	Ile	Leu
		180						185					190		

09965807.100101

Asp Gln Met Arg Lys Met Ile Lys His Ala Leu Asp Phe Ile His His
195 200 205

Phe Asn Glu Gly Lys Glu Phe Pro Pro Cys Ala Ile Glu Val Tyr Lys
210 215 220

Ile Ile Glu Lys Val Asp Tyr Pro Arg Asp Glu Asn Gly Glu Ile Ala
225 230 235 240

Ala Ile Ile His Pro Asn Leu Gln Asp Gln Asp Trp Lys Pro Leu His
245 250 255

Pro Gly Asp Pro Met Phe Leu Thr Leu Asp Gly Lys Thr Ile Pro Leu
260 265 270

Gly Gly Asp Cys Thr Val Tyr Pro Val Phe Val Asn Glu Ala Ala Tyr
275 280 285

Tyr Glu Lys Lys Glu Ala Phe Ala Lys Thr Thr Lys Leu Thr Leu Asn
290 295 300

Ala Lys Ser Ile Arg Cys Cys Leu His
305 310

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 313 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: Region
(B) LOCATION: 6
(D) OTHER INFORMATION: /note= "This is isoleucine in human; valine in bovine. This is a very conservative substitution."

(ix) FEATURE:

- (A) NAME/KEY: Region
(B) LOCATION: 9
(D) OTHER INFORMATION: /note= "This is glutamic acid in human, aspartic acid in bovine. This is a very conservative substitution."

(ix) FEATURE:

- ```
(A) NAME/KEY: Region
(B) LOCATION: 10
(D) OTHER INFORMATION: /note= "This is histidine in human,
 proline in bovine. This is a conservative
 substitution."
```

(ix) FEATURE:

- (A) NAME/KEY: Region  
(B) LOCATION: 12  
(D) OTHER INFORMATION: /note= "This is glutamine in human, lysine in bovine. This is a very conservative substitution."

(ix) FEATURE:

- (A) NAME/KEY: Region  
(B) LOCATION: 38

(D) OTHER INFORMATION: /note= "This is glycine in human, serine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 39

(D) OTHER INFORMATION: /note= "This is alanine in human, threonine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 72

(D) OTHER INFORMATION: /note= "This is isoleucine in human, valine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 75

(D) OTHER INFORMATION: /note= "This is leucine in human, proline in bovine. This is not a conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 82

(D) OTHER INFORMATION: /note= "This is methionine in human, lysine in bovine. This is a conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 134

(D) OTHER INFORMATION: /note= "This is asparagine in human, aspartic acid in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 200

(D) OTHER INFORMATION: /note= "This is lysine in human, glutamine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 208

(D) OTHER INFORMATION: /note= "This is histidine in human, asparagine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 226

(D) OTHER INFORMATION: /note= "This is isoleucine in human, methionine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 227

00965807 100101

(D) OTHER INFORMATION: /note= "This is glutamic acid in human, arginine in bovine. This is not a conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 234

(D) OTHER INFORMATION: /note= "This is aspartic acid in human, asparagine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 236

(D) OTHER INFORMATION: /note= "This is asparagine in human, serine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 240

(D) OTHER INFORMATION: /note= "This is alanine in human, serine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 246

(D) OTHER INFORMATION: /note= "This is asparagine in human, lysine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 258

(D) OTHER INFORMATION: /note= "This is glycine in human, glutamic acid in bovine. This is a conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 261

(D) OTHER INFORMATION: /note= "This is methionine in human, valine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 276

(D) OTHER INFORMATION: /note= "This is cysteine in human, glutamine in bovine. This is not a conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 306

(D) OTHER INFORMATION: /note= "This is lysine in human, asparagine in bovine. This is a very conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 310

09965807 100101

(D) OTHER INFORMATION: /note= "This is cysteine in human.  
serine in bovine. This is a conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: 311

(D) OTHER INFORMATION: /note= "This is cysteine in human.  
serine in bovine. This is a conservative substitution."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 83

(D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 105

(D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 108

(D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 146

(D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 264

(D) OTHER INFORMATION: /note= "Phosphorylation site"

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 117

(D) OTHER INFORMATION: /note= "Potential N-glycosylation site"

(ix) FEATURE:

(A) NAME/KEY: Active-site

(B) LOCATION: 18..24

(D) OTHER INFORMATION: /note= "Consensus sequence  
predicted to be involved in catalysis"

(ix) FEATURE:

(A) NAME/KEY: Active-site

(B) LOCATION: 275..278

(D) OTHER INFORMATION: /note= "Consensus sequence  
predicted to be involved in catalysis"

(ix) FEATURE:

(A) NAME/KEY: Active-site

(B) LOCATION: 283..289

(D) OTHER INFORMATION: /note= "Consensus sequence  
predicted to be involved in catalysis"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Thr Ser Cys His Xaa Ala Glu Xaa Xaa Ile Xaa Lys Val Ala Ile  
1 5 10 15

Phe Gly Gly Thr His Gly Asn Glu Leu Thr Gly Val Phe Leu Val Lys  
20 25 30

00965807 100101



25ix

His Trp Leu<sub>35</sub> Glu Asn Xaa Xaa<sub>40</sub> Glu Ile Gln Arg Thr Gly<sub>45</sub> Leu Glu Val  
 Lys Pro Phe Ile Thr Asn<sub>55</sub> Pro Arg Ala Val Lys<sub>60</sub> Lys Cys Thr Arg Tyr  
 Ile Asp Cys Asp Leu<sub>70</sub> Asn Arg Xaa Phe Asp<sub>75</sub> Xaa Glu Asn Leu Gly<sub>80</sub> Lys  
 Lys Xaa Ser Glu<sub>85</sub> Asp Leu Pro Tyr Glu<sub>90</sub> Val Arg Arg Ala Gln Glu<sub>95</sub> Ile  
 Asn His Leu<sub>100</sub> Phe Gly Pro Lys Asp<sub>105</sub> Ser Glu Asp Ser Tyr<sub>110</sub> Asp Ile Ile  
 Phe Asp Leu<sub>115</sub> His Asn Thr Thr Ser<sub>120</sub> Asn Met Gly Cys Thr<sub>125</sub> Leu Ile Leu  
 Glu Asp<sub>130</sub> Ser Arg Asn Xaa<sub>135</sub> Phe Leu Ile Gln Met Phe<sub>140</sub> His Tyr Ile Lys  
 Thr Ser Leu Ala Pro Leu<sub>150</sub> Pro Cys Tyr Val Tyr<sub>155</sub> Leu Ile Glu His Pro<sub>160</sub>  
 Ser Leu Lys Tyr<sub>165</sub> Ala Thr Thr Arg Ser Ile<sub>170</sub> Ala Lys Tyr Pro<sub>175</sub> Val Gly  
 Ile Glu Val Gly<sub>180</sub> Pro Gln Pro Gln Gly<sub>185</sub> Val Leu Arg Ala Asp<sub>190</sub> Ile Leu  
 Asp Gln Met<sub>195</sub> Arg Lys Met Ile Xaa<sub>200</sub> His Ala Leu Asp Phe<sub>205</sub> Ile His Xaa  
 Phe Asn<sub>210</sub> Glu Gly Lys Glu Phe<sub>215</sub> Pro Pro Cys Ala Ile<sub>220</sub> Glu Val Tyr Lys  
 Ile Xaa Xaa Lys Val<sub>230</sub> Asp Tyr Pro Arg Xaa<sub>235</sub> Glu Xaa Gly Glu Ile Xaa<sub>240</sub>  
 Ala Ile Ile His Pro<sub>245</sub> Xaa Leu Gln Asp Gln<sub>250</sub> Asp Trp Lys Pro Leu<sub>255</sub> His  
 Pro Xaa Asp Pro<sub>260</sub> Xaa Phe Leu Thr Leu<sub>265</sub> Asp Gly Lys Thr Ile<sub>270</sub> Pro Leu  
 Gly Gly Asp<sub>275</sub> Xaa Thr Val Tyr Pro<sub>280</sub> Val Phe Val Asn Glu<sub>285</sub> Ala Ala Tyr  
 Tyr Glu<sub>290</sub> Lys Lys Glu Ala Phe<sub>295</sub> Ala Lys Thr Thr Lys<sub>300</sub> Leu Thr Leu Asn  
 Ala Xaa Ser Ile Arg Xaa<sub>310</sub> Xaa Leu His

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 313 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

09965807 100101



25xi

Ile Asp Cys Asp Leu Asn Arg Val Phe Asp Pro Glu Asn Leu Gly Lys  
65 70 75 80  
Lys Lys Ser Glu Asp Leu Pro Tyr Glu Val Arg Arg Ala Gln Glu Ile  
85 90 95  
Asn His Leu Phe Gly Pro Lys Asp Ser Glu Asp Ser Tyr Asp Ile Ile  
100 105 110  
Phe Asp Leu His Asn Thr Thr Ser Asn Met Gly Cys Thr Leu Ile Leu  
115 120 125  
Glu Asp Ser Arg Asn Asp Phe Leu Ile Gln Met Phe His Tyr Ile Lys  
130 135 140  
Thr Ser Leu Ala Pro Leu Pro Cys Tyr Val Tyr Leu Ile Glu His Pro  
145 150 155 160  
Ser Leu Lys Tyr Ala Thr Thr Arg Ser Ile Ala Lys Tyr Pro Val Gly  
165 170 175  
Ile Glu Val Gly Pro Gln Pro Gln Gly Val Leu Arg Ala Asp Ile Leu  
180 185 190  
Asp Gln Met Arg Lys Met Ile Gln His Ala Leu Asp Phe Ile His Asn  
195 200 205  
Phe Asn Glu Gly Lys Glu Phe Pro Pro Cys Ala Ile Glu Val Tyr Lys  
210 215 220  
Ile Met Arg Lys Val Asp Tyr Pro Arg Asn Glu Ser Gly Glu Ile Ser  
225 230 235 240  
Ala Ile Ile His Pro Lys Leu Gln Asp Gln Asp Trp Lys Pro Leu His  
245 250 255  
Pro Glu Asp Pro Val Phe Leu Thr Leu Asp Gly Lys Thr Ile Pro Leu  
260 265 270  
Gly Gly Asp Gln Thr Val Tyr Pro Val Phe Val Asn Glu Ala Ala Tyr  
275 280 285  
Tyr Glu Lys Lys Glu Ala Phe Ala Lys Thr Thr Lys Leu Thr Leu Asn  
290 295 300  
Ala Asn Ser Ile Arg Ser Ser Leu His  
305 310

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AACCCTACTC TTAAGGAC

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod base= OTHER  
/note= "The M13 universal primer tag is attached to base number 1."

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCGGGATGAA AATGGAGAA

19

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod base= OTHER  
/note= "The M13 reverse primer tag is attached to base 1."

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACCGTGTAAG ATGTAAGC

18

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AGGATCAAGA CTGGAAACC

19

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTAAGACACC GTGTAAGATG

20

09965807 100101

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 7 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Gly Gly Thr His Gly Asn Glu  
 1 5

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 7 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Val Asn Glu Ala Ala Tyr Tyr  
 1 5

## (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 7 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(v) FRAGMENT TYPE: internal

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Val Xaa Glu Xaa Xaa Xaa Tyr  
 1 5

## (2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

25xiv

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Leu Glu Asn Ser Thr Glu Ile Gln Arg Thr Gly Leu Glu Val Lys Pro  
1 5 10 15  
Phe Ile Thr Asn Pro Arg Ala Val Lys Lys  
20 25

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Lys Pro Leu Ile Pro Xaa Asp Pro Val Phe Leu Thr Leu Asp Gly Lys  
1 5 10 15  
Thr Ile Ser Leu Gly Gly Asp Gln Thr Xaa Tyr Pro Xaa Phe Xaa Asn  
20 25 30  
Glu Ala Ala Tyr Tyr  
35

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Xaa Lys Val Asp Tyr Pro Arg Asn Glu Ser Gly Glu Ile Ser Ala Ile  
1 5 10 15  
Ile His Pro Lys Leu Gln Asp Gln  
20

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

09965807 100101

25xv

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Xaa Xaa Xaa Ala Leu Asp Phe Ile Xaa Asn Phe Xaa Glu Xaa Lys Glu  
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 16
- (D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /mod\_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AARGTNGAYT AYCCNNGNAA

20

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /mod\_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TGRTCYTGNA NYTTNGGRTG

20

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 16 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCGTGTACCC AGTGTT

16

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CTTCTGAATT GCAGAAATCA

20

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GTAAGACACC GTGTAAGATG

20

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 23 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION: 6  
 (D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION: 9  
 (D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified\_base

09065807-100101



(B) LOCATION: 18  
(D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:  
(A) NAME/KEY: modified\_base  
(B) LOCATION: 21  
(D) OTHER INFORMATION: /mod\_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGRTANACNG TYTGRTCNC NCC

23

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ix) FEATURE:  
(A) NAME/KEY: modified\_base  
(B) LOCATION: 3  
(D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:  
(A) NAME/KEY: modified\_base  
(B) LOCATION: 6  
(D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:  
(A) NAME/KEY: modified\_base  
(B) LOCATION: 9  
(D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:  
(A) NAME/KEY: modified\_base  
(B) LOCATION: 12  
(D) OTHER INFORMATION: /mod\_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCNMGNGCNG TNAARAARTG

20

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Asp Cys Thr Val

1

09965807-100101

## (2) INFORMATION FOR SEQ ID NO:25:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(v) FRAGMENT TYPE: internal

## (ix) FEATURE:

- (A) NAME/KEY: Active-site
- (B) LOCATION: 1..7
- (D) OTHER INFORMATION: /note= "Consensus sequence of catalytic center in esterases"

## (ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /note= "Amino acid 5 is glycine or alanine"

## (ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /note= "Amino acid 7 is glutamic acid or aspartic acid"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Gly Xaa Xaa His Xaa Xaa Xaa  
 1 5

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(v) FRAGMENT TYPE: internal

## (ix) FEATURE:

- (A) NAME/KEY: Active-site
- (B) LOCATION: 1..4
- (D) OTHER INFORMATION: /note= "Consensus sequence of catalytic center in esterases"

## (ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /note= "Amino acid 4 is phenylalanine or valine"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Asp Xaa Xaa Xaa  
 1

25xix

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(v) FRAGMENT TYPE: internal

(ix) FEATURE:

- (A) NAME/KEY: Active-site
- (B) LOCATION: 1..7
- (D) OTHER INFORMATION: /note= "Consensus sequence of catalytic center in esterases"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Val Xaa Glu Xaa Xaa Tyr  
1 5

0065807.100101